How DNA data storage works

By Graham Templeton (https://www.extremetech.com/author/gtempleton) on July 8, 2016 at 2:00 pm

DNA data storage is a big deal. Partly, it's because we're based on DNA, and any research into manipulation of that molecule will pay dividends for medicine and biology in general — but in part, it's also because the world's most wealthy and powerful corporations are getting discouraged at cost estimates for data storage in the future. Facebook, Apple, Google, the US government, and more are all making astounding investments in storage ("exabyte" is the buzzword now). But even these mega-projects can only put off the inevitable for so long; we are simply producing too much data for magnetic storage to keep up, without a major unforeseen shift in the technology.

That's why a company like Microsoft (http://www.extremetech.com/tag/microsoft) recently decided to invest in the prospect of storing information with a totally different sort of tech: *bio*tech. It might seem off-brand for the software giant, but teaming up with academics to take on molecular biology has produced (http://www.washington.edu/news/2016/04/07/uw-team-stores-digital-images-in-dna-and-retrieves-them-perfectly/) stunning results: The team was able to store and perfectly recall digital data with incredible storage density. According to an accompanying blog post (https://blogs.microsoft.com/next/2016/07/07/microsoft-university-washington-researchers-set-record-dna-storage/#sm.000112un1fb2veysye62g7cp87m6q), they managed to

pack about 200 megabytes of data into just a fraction of a drop of liquid, including a compressed music video from the band *OK Go*. Even more impressive, that data was stored in a quickly and easily accessible form, making it more akin to computer RAM, than computer storage.

So how did they accomplish this incredible feat?

First, they had to convert the digital code of 1's and 0's to a genetic code of A's, C's, T's, and G's, then take this lowly text file and manually construct the molecule it represents. Each of these is a feat in and of itself. DNA storage requires cutting-edge techniques in data compression and security to design a sequence both info-dense enough to realize DNA's potential and redundant enough to allow robust error-checking to improve the accuracy of information retrieved down the line.

(https://www.extremetech.com/wp-content/uploads/2015/11/dna-storage-51.jpg)Very little of the technology on display here is *new,* since the most important parts of the system have existed much longer than mankind itself. But if all the data necessary to code for Albert Einstein was contained within the nucleus of every single cell of Albert Einstein's body, as it was, then this classical approach to data storage must have something going for it. Researchers in this field set out to understand and harness that something, and they're getting better at it seemingly every couple of months.

At the end of the day, DNA's key special attribute it data storage density: how much information can DNA (http://www.extremetech.com/tag/dna) fit into a given unit volume? The NSA's largest, most notorious data-center is an enormous, sprawling complex full of networked racks of magnetic storage drives — but according to some estimates, DNA could take the volume of data contained in about a hundred industrial data centers and store it in a space roughly the size of a shoe box.

DNA achieves this in two ways. One, the coding units are very small, less than half a nanometer to a side, where the transistors of a modern, advanced computer storage drive struggle to beat the 10 nanometer mark. But the increase in storage capacity isn't just ten- or a hundred-fold, but *thousands-fold.* That differential arises from the second big advantage of DNA: it has no problem packing three-dimensionally.

See, transistors are generally aligned on a flat plane, meaning their ability to fully use a given space is pretty low. We can of course stack many such flat boards one atop another, but at that point

(https://www.extremetech.com/wp-content/uploads/2013/03/GENOME-

a new and totally debilitating problem arises: heat. One of the most challenging parts of designing new transistor-based technologies, whether they're processors or storage devices, is heat. The more tightly you pack silicon transistors, the more heat you'll create, and the harder it will be to ferry that heat away from the device. This both limits the maximum density, and requires that we supplement the cost of the drives themselves with expensive cooling systems.

SPEED.jpg.jpg)

Sequencing has gotten much faster and cheaper over time — and that's good, because we need to sequence DNA data to read it!

With its super-efficient packing structure, the DNA double helix offers a great solution. Chromatin, the DNA-protein system that makes up chromosomes, is essentially a very complex mechanism designed to allow an inherently sticky molecule like DNA to roll up really tight, yet still unroll quickly and easily later on, when certain patches of DNA are needed by the body.

(https://www.extremetech.com/wp-content/uploads/2015/11/dna-storage-4.jpg)

Here's a simplified look at how DNA packs so tightly into three-dimensional space.

This at-hand nature of the chromatin system, which allows any gene to be "called" from any part of the genome with roughly equal efficiency, has led the researchers to dub their storage system a DNA version of a computer's random access memory, or RAM. Like RAM, the physical location of a piece of data within the drive isn't important to the computer's ability to access that information.

(https://www.extremetech.com/wp-content/uploads/2016/07/DNA.jpg)However, storing information in DNA differs from computer RAM in some pretty significant ways. Most notable is speed; part of what makes RAM RAM is that its easy-access system is also a *quick* access system, allowing it to hold data the computer might need at an instant's notice, and make it available on those timescales. On the other hand, DNA is significantly harder and slower to read than conventional computer transistors, meaning in terms of access speed it's actually *less* RAM-like than your average computer SSD or spinning magnetic hard-drive.

That's because the incredible abilities of evolution's data storage solution were tailored to evolution's unique needs, and those needs don't necessarily include performing thousands of "reads" per second. Regular, cellular DNA data storage has to untangle the complex chromatin structure of stable DNA, then unwind the DNA double helix itself, make a copy of the sequence of interest, then zip everything right back up the way it was — it takes a while.

For our purposes, we must then add the extra step of reading the DNA. In this case, that's achieved by using an age-old technique in biotech labs called the polymerase chain reaction (PCR) to amplify, or repeatedly duplicate, the sequence we want to read. The whole sample is then sequenced, and everything but the many-many-many-times repeated sequence we amplified is discarded. What remains is our sequence of interest. These stretches of DNA are marked with little target sequences that allow the PCR proteins to bind, and the replication process to begin.

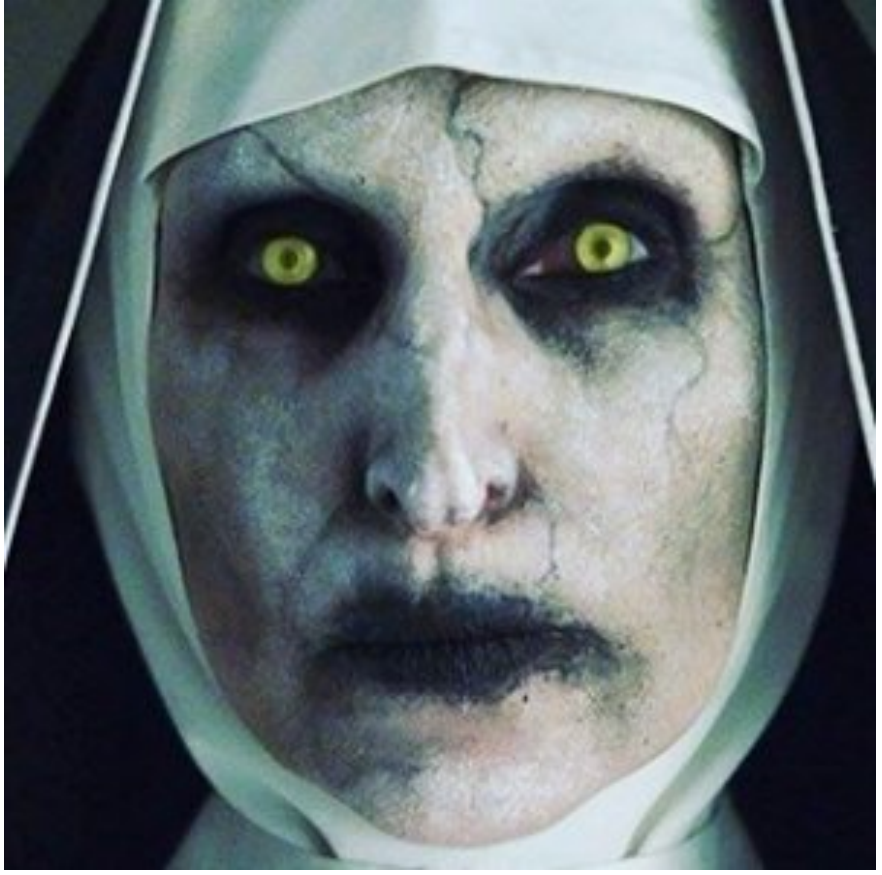
(https://www.extremetech.com/wp-content/uploads/2015/08/gene-therapy.jpg)In cells, genes are turned "on" and "off" largely by changing the availability of these target sequences to the always-waiting machinery of DNA replication. This can be done via the winding and unwinding of chromatin, the direct addition or removal of a blocker protein, or even interaction with other areas of the genome to promote or preclude transcription. In a man-made data storage system, we could theoretically make something better suited to our needs, stronger or more efficient or less wasteful on forms of security we don't need for this purpose, but that would require a level of sophistication in protein engineering that still seem a ways out.

Check out our ExtremeTech Explains (http://www.extremetech.com/tag/extremetech-explains) series for more in-depth coverage of today's hottest tech topics.

Now read: How DNA sequencing works (http://www.extremetech.com/extreme/214647-how-does-dna-sequencing-work)

(http://www.zergnet.com/i/3164138/41518/0/0/0/1)

The Disturbing Truth About 'Dating Naked' (http://www.zergnet.com/i/3164138/41518/0/0/0/1)

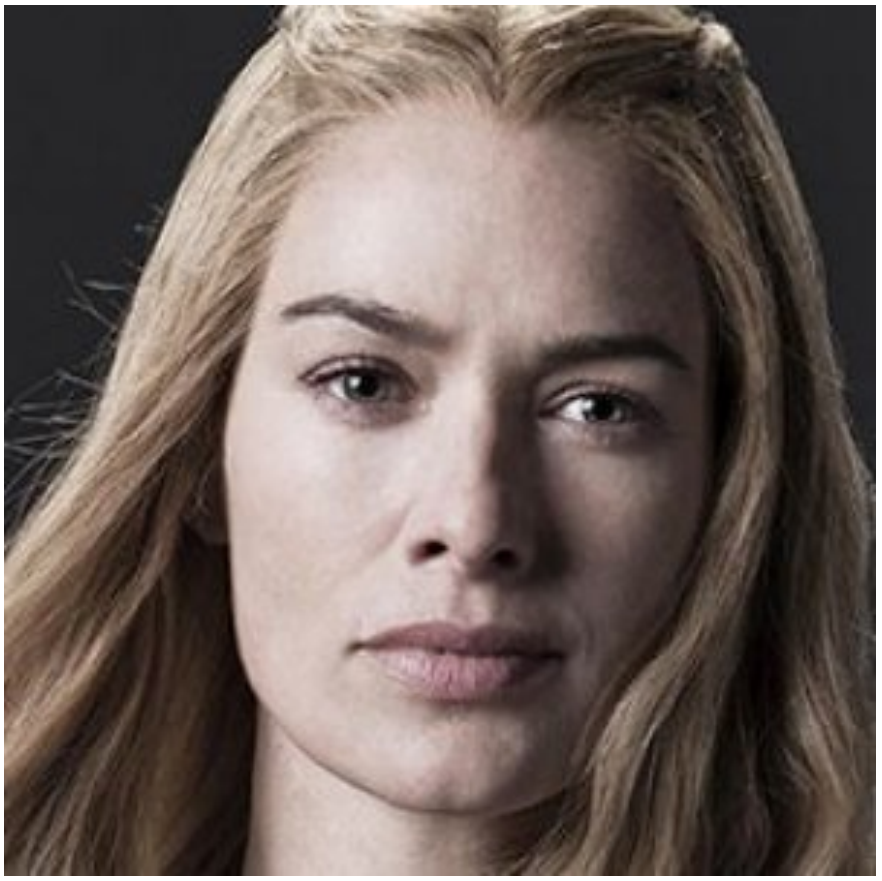


(http://www.zergnet.com/i/3415519/41518/0/0/0/2)

The Actress Who Plays The Nun is Gorgeous in Real Life (http://www.zergnet.com/i/3415519/41518/0/0/0/2)

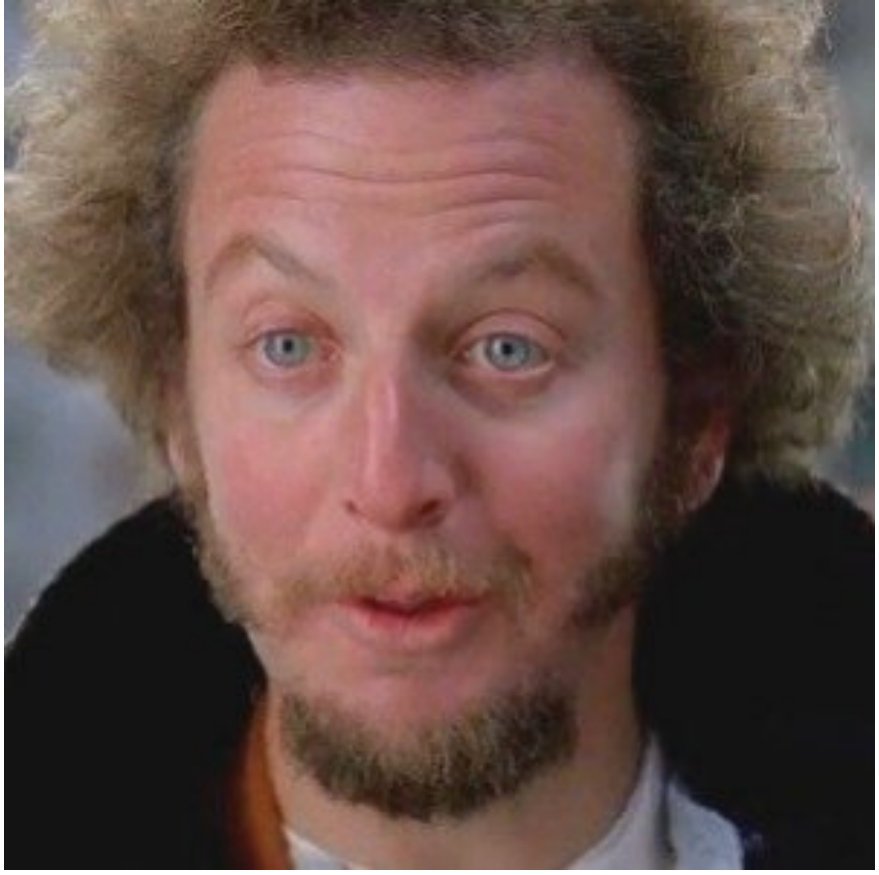
(http://www.zergnet.com/i/3620949/41518/0/0/0/3)

Movie Details That Were Changed for the Most Bizarre Reasons (http://www.zergnet.com/i/3620949/41518/0/0/0/3)



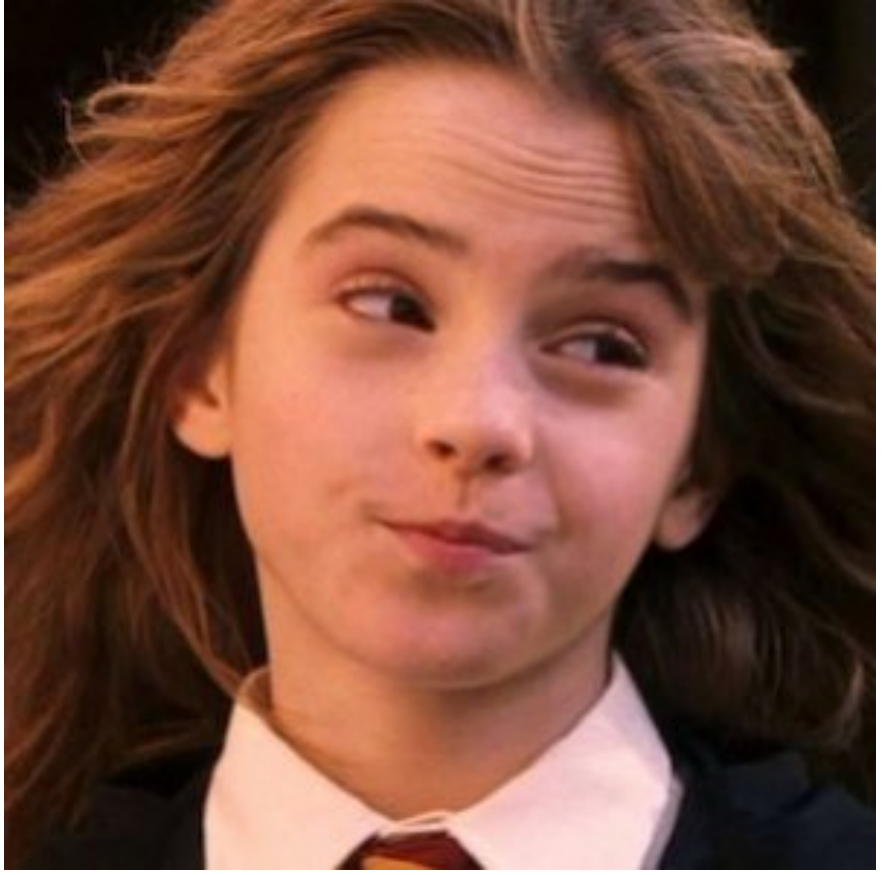
(http://www.zergnet.com/i/3613627/41518/0/0/0/4)

Actors Who Were Replaced and No One Noticed (http://www.zergnet.com/i/3613627/41518/0/0/0/4)



(http://www.zergnet.com/i/3621197/41522/0/0/0/1)

Whatever Happened to This Star from 'Home Alone'? (http://www.zergnet.com/i/3621197/41522/0/0/0/1)



(http://www.zergnet.com/i/3572546/41522/0/0/0/2)

J.K. Rowling Confirms a Hermione Theory We Suspected All Along (http://www.zergnet.com/i/3572546/41522/0/0/0/2)

(http://www.zergnet.com/i/3468818/41522/0/0/0/3)

Surprising Things Men Found Attractive 50 Years Ago (http://www.zergnet.com/i/3468818/41522/0/0/0/3)

(http://www.zergnet.com/i/3591173/41522/0/0/0/4)

What's Come Out About 'Old Man' From 'Pawn Stars' Since He Died (http://www.zergnet.com/i/3591173/41522/0/0/0/4)

Tagged In

microsoft (https://www.extremetech.com/tag/microsoft)
extremetech explains (https://www.extremetech.com/tag/extremetech-explains)
dna (https://www.extremetech.com/tag/dna) transistors (https://www.extremetech.com/tag/transistors)
biotech (https://www.extremetech.com/tag/biotech)
data storage (https://www.extremetech.com/tag/data-storage)
DNA storage (https://www.extremetech.com/tag/dna-storage)
exabytes (https://www.extremetech.com/tag/exabytes)

Post a Comment

17 Comments

David • 2 years ago

this is incredibly amazing advancement in biotech the implications are vast and i hope the limitations met are surmountable.....

1 • Reply • Share ›

Furkan Gözükara • 2 years ago

I cant believe the ignorance of evolution blind believers. we are talking about much advanced and complex technology than what we have and calling it randomly appeared out of non-living material is utter non-sense. believing in random evolution is just another fake religion. yes these believers does not believe in science but they have faith in evolution religion.

our current technology is cumulatively maybe Quadrillions of man-hours and it is still very inferior than the technology of living organisms and we are calling living organisms pure randomness just lol :D

on the subject, i believe DNA will be utilized in future. The Allah's creation is magnificent and marvelous.

• Reply • Share ›

Ah Got Somethin Ta Say! ➚ Furkan Gözükara • 2 years ago

"our current technology is cumulatively maybe Quadrillions of man-hours and it is still very inferior than the technology of living organisms"

Our technology is decades in the making. Living organisms are eons in the making.

2 ^ | v • Reply • Share ›

Furkan Gözükara ➚ Ah Got Somethin Ta Say! • 2 years ago

ye eons of which consciousness maker? developer? or system. and if system who coded or developed that system?

^ | v • Reply • Share ›

Felix ➚ Furkan Gözükara • 2 years ago

System, and it's coded by the organisms themselves. Those who were better at coding got to transmit the results and process to the next generation. This is not really a question any of the scientists involved are asking. They know how to develop storage media because they have that understanding nailed down.

^ | v • Reply • Share ›

captainwiggins ➚ Felix • a year ago

An organism can't see it's DNA let alone code it.

^ | v • Reply • Share ›

Dickson ➚ Furkan Gözükara • 2 years ago

The beauty of science is that it is true, regardless of whether or not you believe in it. Evolution is a fact of life, and many of our modern industries are reliant upon the understanding of it, particularly medicine. Even free market capitalism for the most part acts a lot like natural selection. There is far more evidence for Evolution than there is for ANY creator, including Allah.

^ | v • Reply • Share ›

This comment is awaiting moderation. Show comment.

Dickson ➚ Paul Celauro • a year ago

Ridiculous waffle there, with unproven assumptions. Start by justifying your claim that "An
analog entity per se is incapable of digitizing the code for ANYTHING, without a pre-existent A/D converter - although the reverse is completely possible."

^ | v • Reply • Share ›

Paul Celauro ➚ Dickson • a year ago

Hey Dickson, I'm starting my website in a few weeks - TrueUniverse.net. I have a large group of leading technologists, engineers, medical doctors, scientists, innovators, attorneys, and

theologians as advising directors - they comprise our Research Group. They are the ones that pushed me into this - I think you could help us greatly by becoming an advisor help us to develop better function models and better ways to communicate our assertions to people who basically disagree with us on principle. If you send your email address to researchgroup@trueuniverse.net we'll give you a free membership that will allow you to appear in our blog and openly state your positions.

This Extreme Tech site is absolute dynamite. This phenomenal DNA article is what got me here and I've got to say that the fact that the DNA molecule itself embodies more functions and hooks both in the Analog Physical Universe and the Quantum Digital Universe - it is truly a bi-directional high speed wireless distributed process control system and reality transducer the likes of which exist nowhere in industry. If

see more

Reply • Share ›

Jeff Vahrenkamp • 2 years ago

in the cell the DNA is the slow storage like a HDD (maybe even magnetic tape slow if we're talking Eukaryota cells) and the RNA is the RAM. Transcription is like reading data from the hard drive into RAM where it can accessed very quickly over and over again. Transfer rates are abominable for this (60~100 bits a second for transcription or 2-5 minutes for an average gene), but it does run in a massively parallel way to over come this. Cells are pretty amazing computer systems, and we could definitely take some tips on making small, extremely efficient computers by studying their design.

Reply • Share ›

WeThePeopleUS • 2 years ago

My plan and technologies are progressing slowly. Soon, we'll be cyborgs with 500 year + lifespans and infinite memory and intelligence.

Reply • Share ›

Mr Opinionated ➚ WeThePeopleUS • 2 years ago

We will all have 4096 DNA based brains feeding an AMD GPU with a total data bandwidth in the terrabytes, and we will then invent FASTDNA and destroy the universe.

Reply • Share ›

Chris Daly • 2 years ago

Important problem.

PCR uses 18-30 bases as a primer (target, tag), That's 36 to 60 bits (4 to 7 1/2 bytes).

Any combination of 00 ,01 ,10 ,11 (A,T,C,G) is likely to be encountered in a byte stream.

So, it is extremely likely that the PCR protein will accidentally bind to the DNA in the wrong place for certain sequences.

Unless the codon attached is changed based on the presence of certain bit-pair combinations. As a simple example, 00 01 10 11 might have 00 00 00 attached as a PCR delimiter, whereas 11 01 11 00 00 00 11 10 might have 11 11 11 attached.

But this tightens the time/accuracy tradeoff threshold.

What I mean is, the more complex the way we interact with DNA, by varying segments and attaching unique, non-present combinations as PCR tags, the longer it will take to process any information stored in the DNA.

I would be most interested to see if we start developing unique 'nucleotides.'

see more

Reply • Share ›

RonG ➚ Chris Daly • 2 years ago

We already *have* developed several unnatural nucleotides, which base pair with each other or with natural nucleotides.

As for "mispriming" in PCR, it is a problem that biologists deal with all the time. There are always unexpected PCR products when amplifying regions from a very complex genome. Some ways to deal with the problem include "nested PCR", which uses a second set of primers within the desired target region, or simply selecting for the product of the correct size. Adding tags to the primers won't help (although the tags are useful for other purposes), but more specificity can be gained by using longer primers with more stringent conditions for binding. Still, a complete fix does not currently exist.

Reply • Share ›

Paul Celauro • a year ago

This is the BEST, MOST IRREFUTABLE, DOWNRIGHT EXCELLENT ARTICLE on DNA I have ever read. Add to this the fact that human DNA simultaneously controls more than 1800 ANALOG chemical reactions in EACH cell in your body, AND the BIGGEST BADDIE for non-creationals - DNA is also bi-directionally reactive with background microwave radiation similarly to smart phones in cell networks. So while seculars see the universe as being historic - like an oil painting, we at TrueUniverse (TM) see it as continously sustained - like a 3-D HDTV screen - where atoms are pixels of reality.

We are weeks away from launching our educational website for High School-ers and

Collegians of all ages TrueUniverse.net and you'd better believe we're going to have links to this article and this site. We debunk the unproductive Science vs the Bible argument by demonstrating that the Analog Physical Universe is a High-Tech Ultra-Large Scale, Application-Specific, Top-Down Integrated System using irrefutable modern technological models. Genesis Ch 1 & 2 (check Genesis - Mechanical Hebrew - Jeff Benner) perfectly describe the materialization and integration of the Analog Physical Universe we live in.

Had Einstein been a Millennial, he would have figured it all out in about 10 minutes - it's just that he never really understood Digital to Analog and Analog to Digital Interfaces and MACHINE CODE.

Phenomenal work Graham, keep it coming!!!

Reply • Share ›

willie raymond mason • a year ago

The ET's already have biological computers.

Reply • Share ›

Yogime • 9 months ago

Incredible content. Just came through this video https://youtu.be/moOJBZqVyLQ
And searched for further. Awesome content. Thanks!!

Reply • Share ›

This site may earn affiliate commissions from the links on this page. **Terms of use (https://www.ziffdavis.com/terms-of-use#endorsement)**.

ExtremeTech Newsletter

Subscribe Today to get the latest ExtremeTech news delivered right to your inbox.

Email Address... Sign Up

This newsletter may contain advertising, deals, or affiliate links. Subscribing to a newsletter indicates your consent to our Terms of Use (http://www.ziffdavis.com/terms-of-use) and Privacy Policy (http://www.ziffdavis.com/privacy-policy). You may unsubscribe from the newsletter at any time.

More Articles

(https://www.extremetech.com/computing/2
google-denies-crippling-edge-to-chromes-
advantage)
Google Denies Crippling Edge to Chrome's Adva
(https://www.extremetech.com/computing/28268
google-denies-crippling-edge-to-chromes-advant
Dec 20

(https://www.extremetech.com/computing/2
intel-announces-major-manufacturing-expa
Intel Announces Major Manufacturing Expansion
(https://www.extremetech.com/computing/28270
announces-major-manufacturing-expansion) Dec

(https://www.extremetech.com/gaming/282
modders-are-using-ai-to-overhaul-old-game
textures-with-gorgeous-results)
Modders Are Using AI to Overhaul Old Games Textures, With Gorgeous Results
(https://www.extremetech.com/gaming/282695-
modders-are-using-ai-to-overhaul-old-games-text
with-gorgeous-results) Dec 20

(https://www.extremetech.com/extreme/28
extremetechs-best-cars-and-car-tech-for-2

ExtremeTech's Best Cars and Car Tech for 2019 (https://www.extremetech.com/extreme/282494-extremetechs-best-cars-and-car-tech-for-2019) D

(https://www.extremetech.com/internet/282 facebook-is-probably-tracking-your-locatior even-if-you-tried-to-stop-it)
Facebook Is Probably Tracking Your Location Eve You Tried to Stop It (https://www.extremetech.com/internet/282678-facebook-is-probably-tracking-your-location-even you-tried-to-stop-it) Dec 19